

Predicting habitat to optimize sampling of Pacific sardine (*Sardinops sagax*)

Juan P. Zwolinski^{1*}, Robert L. Emmett², and David A. Demer¹

¹Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 8604 La Jolla Shores Drive, La Jolla, CA 92037, USA

²Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Boulevard East, Seattle, WA 98112, USA

*Corresponding Author: tel: +1 858 5465654; fax: +1 858 5465656; e-mail: juan.zwolinski@noaa.gov.

Zwolinski, J. P., Emmett, R. L., and Demer, D. A. 2011. Predicting habitat to optimize sampling of Pacific sardine (*Sardinops sagax*). – ICES Journal of Marine Science, 68: 867–879.

Received 9 July 2010; accepted 24 February 2011

More than 40 years after the collapse of the fishery for Pacific sardine, a renewed fishery has emerged off the west coasts of the United States and Canada. The daily egg production method (DEPM) and acoustic-trawl surveys are performed annually and, to minimize the uncertainties in the estimates, sampling effort needs to be allocated optimally. Here, based on a 12-year dataset including the presence/absence of sardine eggs and concomitant remotely sensed oceanographic variables, a probabilistic generalized additive model is developed to predict spatio-temporal distributions of habitat for the northern stock of Pacific sardine in the California Current. Significant relationships are identified between eggs and sea surface temperature, chlorophyll *a* concentration, and the gradient of sea surface altitude. The model accurately predicts the habitat and seasonal migration pattern of sardine, irrespective of spawning condition. The predictions of potential habitat are validated extensively by fishery landings and net-sample data from the northeast Pacific. The predicted habitat can be used to optimize the time and location of the DEPM, acoustic-trawl, and aerial surveys of sardine. The method developed and illustrated may be applicable too to studies of other stocks of sardine and other epipelagic fish in other eastern boundary, upwelling regions.

Keywords: generalized additive model, habitat, optimal sampling, Pacific sardine, remote sensing.

Introduction

The Pacific sardine (*Sardinops sagax*) is a coastal pelagic species found along the northeastern Pacific margin, from the Gulf of California to the Gulf of Alaska. In the 1930s, its fishery in the California Current (CC) was the largest single-species one in the western hemisphere, but it collapsed in the 1950s following a dramatic and partly unexplained decrease in biomass (Radovich, 1982). In the 1990s, after several successive years of increasing biomass (Smith and Moser, 2003), the Pacific sardine fishery resumed in the CC off the United States and Canada (Figure 1). The US Pacific sardine fishery is currently regulated by the Pacific Fishery Management Council (PFMC), which sets harvest guidelines based on annual US assessments of the northern or cold stock of Pacific sardine, hereafter sardine. The assessments were traditionally based on catch-at-age information and fishery-independent estimates of biomass provided by daily egg production method (DEPM) surveys and, more recently, from aerial surveys (Hill *et al.*, 2009). Motivated by differing information from these various sources, the PFMC requested that data from fishery-independent surveys using combined echosounder and trawl sampling (acoustic-trawl surveys) be included in future assessments (Hill *et al.*, 2006). Planning effective assessment surveys requires the judicious selection of survey timing, area, and effort to maximize the likelihood of (i) sampling the entire

northern stock, and (ii) providing the estimates of sardine abundance, with low sampling variance. These two goals compete in the use of limited ship and aircraft time, and the balance needs to be optimized systematically.

As stated above, this study focuses solely on the northern stock of sardine (Smith, 2005), which exhibits large seasonal movements within the CC out to ~300 nautical miles (hereafter miles) offshore (Macewicz and Abramenkoff, 1993) between northern Baja California, Mexico, and Canada (Figure 1; Clark and Janssen, 1945), apparently following habitat characterized by a specific range of seawater temperature (Lluch-Belda *et al.*, 1991; Ware, 1999; Lo *et al.*, 2007) and primary production (Reiss *et al.*, 2008). In spring, sardine concentrate off central and southern California for their primary spawning season (Clark and Janssen, 1945). As the season progresses and sea surface temperature (SST) increases, adult sardine migrate north and inshore (McClatchie, 2009). There is also sporadic spawning during late spring and summer off Oregon and Washington (Emmett *et al.*, 2005), and occasionally off Canada (McFarlane *et al.*, 2005). Sardine aggregate off southern Vancouver Island, British Columbia (BC), Canada, in summer (Ware, 1999). In winter, the greater part of the northern sardine stock retreats south, concentrating again off southern California for the next spring spawning.

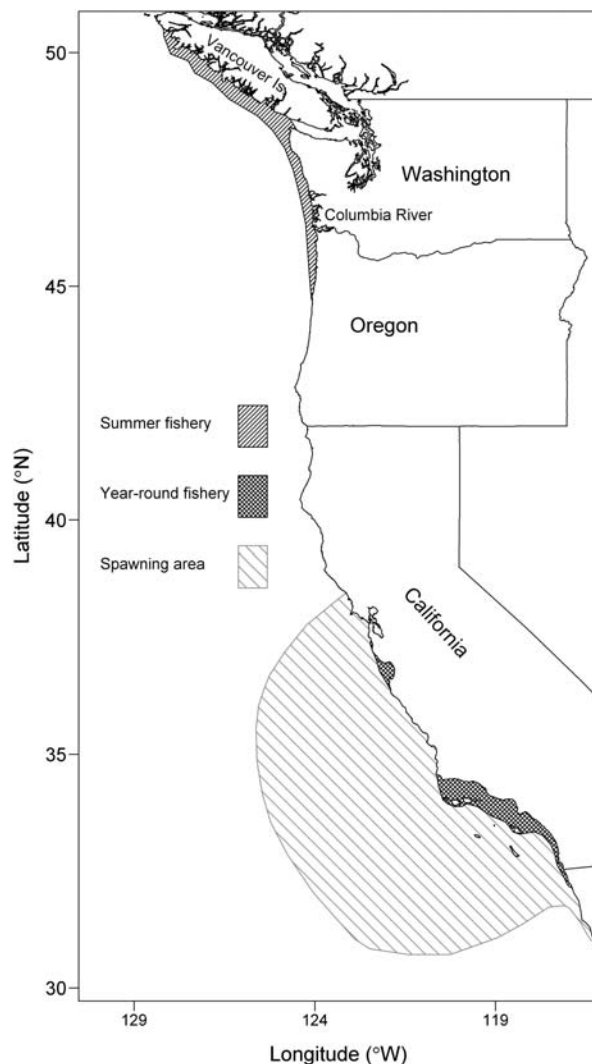


Figure 1. The CC region, indicating typical locations of spring-spawning sardine, the summer fishery off Oregon, Washington, and Vancouver Island, and the year-round fishery off northern Baja California and California.

Most of what is known about sardine movements along the northeastern Pacific margin dates back to the 1940s, before the collapse of the fishery, and is derived from the fishery itself and large-scale tagging studies (Clark and Janssen, 1945). At that time, sardine were harvested in the CC from Baja California to BC. Recent information on sardine migration is limited to the time and area of spring spawning (Lo *et al.*, 2007), and when and where the fishery is carried out in summer (Figure 1; Emmett *et al.*, 2005; McFarlane *et al.*, 2005). Nevertheless, the early observed pattern of sardine migration appears to persist, perhaps predictably. A model to predict the spatio-temporal distribution of sardine off the entire west coast of the United States throughout the year would be extremely valuable in helping direct and optimize survey activities.

Many studies have concluded that sardine tend to inhabit waters within restricted ranges of temperature. During spring, sardine from the northern stock spawn off southern California and Baja California in waters of 12–16°C (Checkley *et al.*, 2000;

Lynn, 2003; Jacobson *et al.*, 2005; Reiss *et al.*, 2008), generally around 13°C (Reiss *et al.*, 2008; Weber and McClatchie, 2010). To the north, sardine generally appear during summer and generally avoid water <12°C (Schweigert and McFarlane, 2001). When there is spawning there, it is associated with temperatures similar to those observed during spring off southern California (Emmett *et al.*, 2005). These findings support a hypothesis that the northern stock of Pacific sardine move within a restricted and likely identifiable habitat, probably defined by a combination of favourable water temperature and feeding opportunities, perhaps proxies for other biophysical forcing functions.

Satellite-sensed and vessel-sampled oceanographic conditions have been used to adapt survey effort in real time (Checkley *et al.*, 2000). However, a habitat model based on remotely sensed information alone has not been developed, tested extensively, or used operationally. Here, using more than a decade of observations, a model is developed to characterize the relationships between sardine occurrence and the epipelagic environment. The model is used to predict spatio-temporal variations in the habitat of the northern stock of Pacific sardine and to predict their seasonal migrations in the CC. Such predictions, future and retrospective, will serve to reduce both random and systematic components of error in the survey estimates of sardine distribution and abundance, and hence potentially improve management. They will also enhance the general knowledge of the dynamics and interactions of sardine, their prey, and their predators, including fishers.

Methods

Data collection

From 1998 through 2009, distributions and densities of sardine eggs were estimated from spring (generally April) surveys conducted for the DEPM (Hill *et al.*, 2009, and references therein). These surveys targeted the spring spawning area in southern and central Californian waters, mostly between 30 and 37°N and 117 and 126°W (Figure 1). However, the sampling design changed from year to year and extended farther north in some years, occasionally using two vessels within a constrained period, usually less than a month. In 2006 and 2008, for example, the surveys spanned the entire west coast of North America, from the US–Mexico border to BC (Figure 2).

During DEPM surveys, eggs are collected with a continuous, underway fish–egg sampler (CUFES; Checkley *et al.*, 2000), which pumps water from a depth of 3 m and collects the plankton in a sieve of 200-μm mesh. The mean CUFES sampling interval (CSI) was 30 min, corresponding to 5–6 miles at a ship's speed of 10–12 knots. In areas of abundant sardine eggs, accounting for 20% of the samples, the CSI was <20 min. Individual fish eggs were visually classified to species and counted. CUFES samples were spatially indexed by their central geographic positions.

Satellite data were retrieved from the CoastWatch repository (<http://coastwatch.pfeg.noaa.gov/coastwatch/CWBrowser.jsp> and [/erddap/griddap/index.html](http://coastwatch.pfeg.noaa.gov/erddap/griddap/index.html)) in monthly, 14-d, and 7-d composite images. These were used alone, for monthly images, or in combination when using 14- or 7-d images, to maximize the temporal match between remotely sensed variables and the CUFES data.

SSTs (°C) are from the Pathfinder V5.0 sensor (day and night images; 4.4 km resolution) for the period 1998–2008 and from

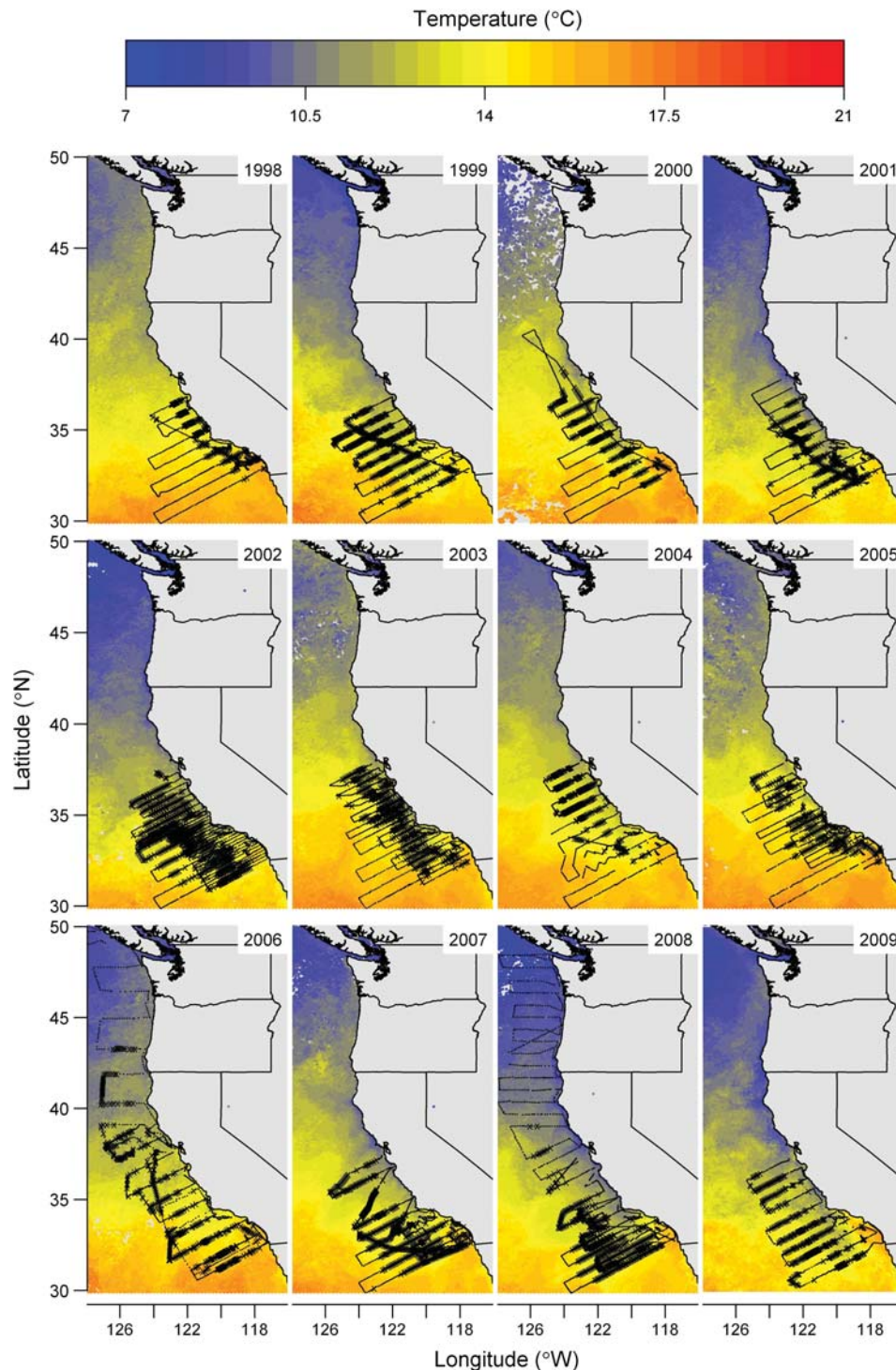


Figure 2. Sardine egg presence (crosses) and absence (dots), sampled by a CUFES. The underlying contour maps show satellite-sensed SST for each spring survey, 1998–2009.

the blended-sea-surface product (weighted mean of several platforms, 0.1° resolution) for 2009. Sea surface concentrations of chlorophyll *a* (CHL; mg m^{-3}) were obtained from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS; 0.04167° resolution) for the period 1998–2002 and from the AquaModis NPP (0.05° resolution) for the period 2003–2009. Sea surface altitude deviation (ALT; m) were obtained from the Archiving, Validation, and Interpretation of Satellite Oceanographic

(AVISO; 0.25° resolution) data archive and used to construct maps of sea-surface-height gradients (GRAD; $\text{m grid_spacing}^{-1}$), i.e. the maximum rate of change in ALT estimated from a 3×3 cell moving plane. This is the minimum sampling neighbourhood required to estimate an isotropic gradient with a resolution comparable with that of CUFES samples.

To match the egg and the environmental data spatially, SST and CHL values within 2 miles of each CUFES sample were averaged

arithmetically. Owing to their lower resolution, ALT and GRAD values within 7.7 miles of each CUFES sample were averaged. All analyses were performed in R V2.10 (R Development Core Team, 2009; see Supplementary material).

Data analysis

A habitat model, following the terminology in Guisan and Zimmermann (2000), was constructed by regressing the presence of sardine eggs on SST, CHL, ALT, and GRAD. As sardine eggs have limited longevity (ca. 3 d at 15°C; Lasker, 1964), their presence or absence in the CUFES samples, coded 1 or 0, respectively, were used to identify the environmental conditions where spawning was more frequent. When fitted to a large set of spatially and temporally independent observations of their environment, the model probabilities define the potential habitat of spawning sardine from the northern stock. Peak spawning of sardine off southern California is in April, with ca. 12% of all mature female sardine spawning each day (Hill *et al.*, 2009). Assuming that the daily proportion of spawning mature females does not exhibit mesoscale (tens to hundreds of kilometres) segregation from the mature post-spawning or recovering fish, i.e. between two spawning events during the same season, the potential habitat of spawning sardine may also be representative of the potential habitat of the greater population of mature sardine. Therefore, the model should also predict the geographic limits of the habitat against time for the entire northern stock of adult sardine in the CC, here restricted to 30–50°N.

The associations between satellite-sensed environmental variables and sardine eggs in CUFES samples were estimated by generalized additive models (GAMs; Wood, 2006). GAMs are a generalization of ordinary linear models, and the linear predictors are related to the response variables via a link function that extends the use of the regression models beyond non-Gaussian response variables. GAMs use data-driven functions, such as splines and local regression, which have superior performance relative to the polynomial functions used in linear models. They allow the depiction of complex relationships between species and their environment, such as those representing species habitats (Guisan and Zimmermann, 2000).

Alternative GAMs were fitted using SST, ln(CHL), GRAD, CSI, and egg presence/absence, a binomial error-distribution, and a logit-link function (mgcv package in R, V1.6.1; Wood, 2006). CHL data were ln-transformed to reduce skew. Single smoothers were fitted using the default spline basis. Interactions between the predictors were fitted using tensor-product smooth functions. These techniques allowed the fitting of variables measured at different scales within the same smoother (Wood, 2006). Model selection was performed using both forward and backward inclusions of variables (see guidelines in Wood and Augustin, 2002, and Wood, 2006). This selection process includes the approximate *p*-values of the smoothers, the statistics representing the compromise between quality-of-fit and complexity of the models, e.g. the generalized cross-validation score or Akaike information criterion (AIC), and visual inspection of the smoothers. As it is not possible to use the per cent of deviance explained in binary models (McCullagh and Nelder, 1997), the area under the receiver operating characteristic (ROC) curve (AUC) was used as a measure of global fit. For a random pair of presence and absence samples, the AUC statistic represents the probability that the model fits a higher probability to the presence sample (Fawcett, 2006). An AUC value of 1 indicates a model with optimal

prediction ability, i.e. a model able to distinguish between all zeroes and values of 1, whereas a value of 0.5 represents a model that predicts no better than randomly.

Habitat models of the pelagic environment evaluate relationships of a presumed temporarily and spatially invariant process, so cannot include temporal or spatial explanatory variables. Moreover, because the models are based on relationships observed with data from surveys with different sample sizes, the explanatory power of a model must be checked for consistency throughout the surveys, to ensure that one or a small number of surveys does not dominate the relationships. Using a common model, i.e. a single model fit with all data, the annual AUC values provide such diagnostics. Large and consistent AUC values suggest a model with accurate and time-invariant prediction capabilities. Ultimately, the habitat model was tested for prediction power through an iterative process that excluded from model fitting one survey at a time and used it as a testing set. The mean predictive AUC values and their coefficients of variation (CVs) arising from this iterative process were used as additional criteria in the evaluation of the candidate models. Density-dependent effects on the habitat selection were tested by (i) including the estimated spawning-stock biomass (SSB; Hill *et al.*, 2009) as a term in the model, and (ii) evaluating the correlation between AUC and SSB in models in which SSB was not included as a predictor. Either significance in the SSB term or a negative correlation between AUC and SSB would indicate that an increase in stock biomass might expand a species' distribution beyond optimal habitat (MacCall, 2009).

CUFES data continuously recorded along transects may be autocorrelated because of the aggregating behaviour of spawning sardine. Autocorrelation in the response variable could cause positive bias in the perceived explanatory power of a model (Diniz-Filho *et al.*, 2003), and hence an overstatement of a model's prediction abilities. To mitigate this potential effect, the analysis included every other CUFES sample. This decimation resulted in a sample separation of at least 10 miles, a distance greater than the correlation dimension of fish schools and sardine-egg patches in the CC (Smith, 1978; Curtis, 2004) and in other eastern boundary current ecosystems (e.g. Zwolinski *et al.*, 2006). Finally, to exclude the possibility of fitting a model for which the explanatory power arises from the presence of spatial correlation in the data rather than from a true distribution-to-environment relationship (Diniz-Filho *et al.*, 2003), as well as to evaluate the lack of important explanatory variables (Barry and Elith, 2006), the models' residuals were tested for autocorrelation using variography (Chilès and Delfiner, 1999).

The utility of the habitat model for predicting sardine distribution and hence optimizing sampling in acoustic and egg surveys was explored by analysing the shape of the ROC (Hanley and McNeil, 1982; verification package in R, V 1.31). For a binary classifier, a ROC curve is a plot of the model's sensitivity (the proportion of positive outcomes whose fitted probabilities are above a certain discriminant threshold: true positives) vs. the model's 1-specificity (the proportion of negative outcomes below the same discriminant threshold: true negatives). A model allows total separation between classes if the fitted probabilities for all positive samples are greater than the probabilities of all the negative samples. In other words, there is at least one probability threshold that indicates a sensitivity of 1 and a 1-specificity of 0.

Four isolevels of habitat probabilities, in decreasing order, were chosen based on the selected GAM, with all years included. These

potential habitat classes include: optimal, greatest probabilities, including 80% of positive samples; good, including 10% of the total positive samples not included in the optimal habitat; bad, including 9% of the total positive samples not included in the optimal and good habitat; and unsuitable, including 1% of the total positive samples not included in the other classes. These classes identify regions of sardine habitat both retrospectively, for evaluation of existing data, and prognostically, for optimizing future sampling effort.

Using the best GAM and monthly averages of the satellite-sensed environmental variables, sardine habitat was predicted monthly from 1998 through 2009 from northern Baja California to southern Vancouver Island, revealing its spatio-temporal dynamics forced by mesoscale oceanic circulation. These predictions of sardine habitat were then compared with independent observations of sardine abundance, including data from net samples taken near the Columbia River and commercial landings from the fisheries off Oregon, Washington, and Vancouver Island, more than 1000 km north of the spawning area off southern California (Figure 1). Further, the coherence between the fisheries and the predicted habitat was tested by spectral analysis of their time-series from 2000 through 2009.

Results

Positive egg samples were contiguously distributed (Figure 2) and more frequent in areas with SST values of 11.5–15.5°C and CHL values of 0.18–3.2 mg m⁻³ (Figure 3). Higher rates of egg encounters with higher ALT and GRAD values were also observed, although less clearly (Figure 3). Low SST (<11.5°C) in conjunction with high CHL (>3.2 mg m⁻³) is indicative of freshly upwelled water and defines the inshore limit of sardine eggs. Offshore, the presence of eggs is bounded by oceanic water, which is warmer (typically >15.5°C) and low in CHL (<0.18 mg m⁻³).

There were significant relationships between egg presence and each of the environmental variables, but the relationships with SST and ln(CHL) were considerably stronger than those with ALT and GRAD (see Supplementary material). Also, models consisting of combinations of the explanatory variables were more significant, leading to a final model derived from a combination of SST, ln(CHL), GRAD, and CSI:

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = \text{te}(\text{SST}, \ln(\text{CHL})) + \text{te}(\text{GRAD}, \ln(\text{CHL})) + s(\text{CSI}), \quad (1)$$

where \hat{p} is the fitted probability, and *te* and *s* represent, respectively, a tensor-product smoother and a single-dimensional smoother, both using the default thin-plate basis available in the *mgcv* package (see Supplementary material). The selected model had an average explanatory AUC of 0.83, a maximum of 0.91 for the 2008 survey, and a minimum of 0.69 in 2006 (Table 1). The semivariograms of this model's residuals showed little, if any, spatial structure (results not shown).

The predictions from the sequential models, i.e. models using all data except the surveys used for validation, consistently provided AUC values >0.5 and with high statistical significance (Table 1). The maximum and the minimum AUC values were 0.88 in 2008 and 0.62 in 2006, respectively. There were no indications of density-dependent effects in habitat selection, because the SSB term was non-significant when included in the model, and the AUC and SSB values were not significantly correlated (results not shown). These results suggest that the model, as described in Equation (1), captured significant and temporally invariant relationships between spring-spawning sardine and their habitat. The selected model fitted a maximum probability of sardine presence at approximately SST = 13.2°C,

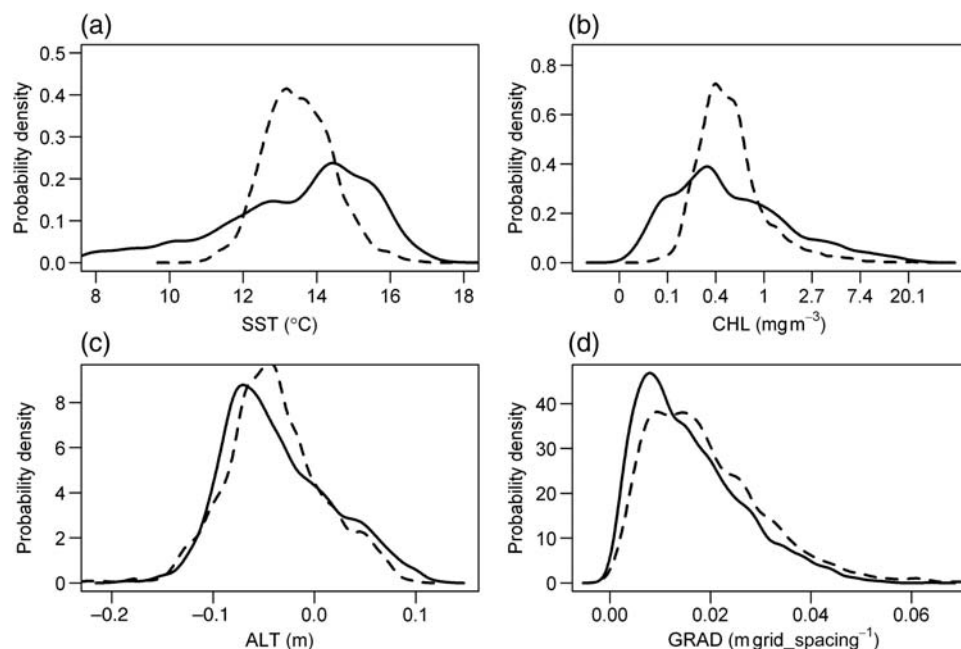


Figure 3. Empirical probability density functions for CUFES samples with (dashed line) and without (continuous line) sardine eggs plotted against satellite-sensed (a) SST, (b) ln(CHL), (c) ALT, and (d) GRAD. The y-axis labels in (b) are linear.

$\ln(\text{CHL}) = -0.7$ ($\text{CHL} = 0.5 \text{ mg m}^{-3}$), and $\text{GRAD} > 0.05 \text{ m grid_spacing}^{-1}$ (Figure 4).

Examination of the ROC curve for the global model [Equation (1)] resulted in probability thresholds of 0.05, 0.32, and 0.45 for delimiting the lower boundaries of the bad, good, and optimal habitats, respectively (Figure 5). Using these limits, possible reduction in survey effort was evaluated for each CUFES survey. If the sampling area was chosen conservatively to exclude unsuitable habitat, an average of 13% of the original sampling effort could have been reallocated while still sampling >99% of all the eggs originally sampled. If sampling was restricted to good and optimal habitat, 92% of the eggs could be sampled, saving 36% of the total original survey effort, on average. In some cases, the saved effort could exceed 50% of total survey time (see Supplementary material). These gains reflect the model’s power to predict the inshore, southern, and offshore limits of sardine habitat accurately (Figure 6). In many years, the model indicated that sardine habitat extended farther offshore and north than survey effort. More optimal allocation of the survey effort could have allowed those areas to be sampled without increasing total survey time. However, it must be recognized that the model predicts the potential habitat for very patchily distributed sardine and that not all the predicted habitat will contain sardine. In fact, from our definition of habitat, it follows that good and optimal habitats contain an average of 42% of all negative samples.

Mapping sardine habitat monthly from 1998 to 2009 illustrates cyclic seasonal evolution (Figure 7). From January through April, the habitat is offshore and south of Oregon. Starting in April, the habitat begins to move north while compressing towards the coast. This pattern continues throughout summer, creating a narrow

band of habitat along the coast from Oregon to BC. In July and August, sardine habitat is generally compressed along the coast by the warm, oligotrophic water offshore and to the south. In early autumn, sardine habitat starts to expand and, from October through November, it begins to recede towards the south, retreating first from the coast off BC. The offshore–southward movement continues through winter, and reaches its southern extreme during February and March.

The model predictions of potential sardine habitat were compared with sardine landings in Oregon, Washington, and southern Vancouver Island (Figure 8). The cross-spectra of potential habitats and the corresponding landings show the greatest coherence around a 12-month period, as expected from the largely repeating seasonal pattern. However, potential sardine habitat precedes the catches, more so to the north. Generally, sardine catches off Oregon and Washington begin within the same month and precede the BC landings by ~1 month, reflecting the northward movement of sardine. Fishing off Washington and Oregon ends before the sardine habitat retreats from those areas, but fishing continues for another month or two off BC, suggesting that the fish do not return south via the same coastal route. This hypothesis is supported by the model predictions, which show that the potential habitat is typically located offshore during the southward recessions (Figure 7).

The model predictions were also compared with the sizes and abundances of sardine aged 1 year and more collected weekly in surface trawls during spring and summer near the Columbia River mouth (Figure 9). Those catches provided information on the arrival of adult sardine migrating from the south. Catches of large sardine are made at or shortly after the development of

Table 1. Explanatory and predictive areas under the ROC curve (AUC) values, calculated with the best model derived with and without the predicted dataset, respectively.

AUC	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Mean
Explanatory	0.92	0.78	0.83	0.83	0.76	0.87	0.89	0.85	0.69	0.84	0.91	0.78	0.83
Predictive	0.88	0.67	0.81	0.79	0.72	0.85	0.87	0.82	0.62	0.83	0.88	0.74	0.79

All AUC values had $p < 0.001$.

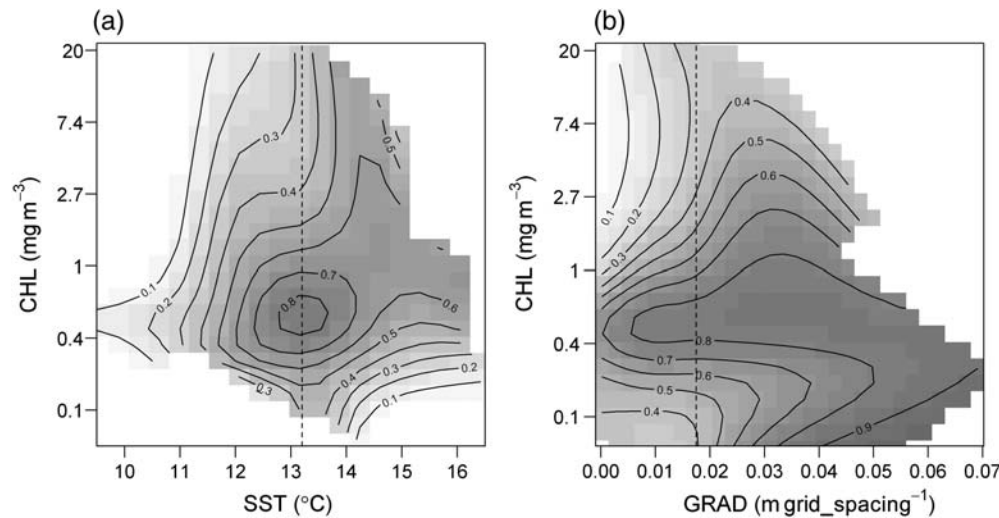


Figure 4. The partial effects for SST, $\ln(\text{CHL})$, and GRAD for the best GAM. The SST vs. $\ln(\text{CHL})$ contour (a) was conditioned on $\text{GRAD} = 0.0175 \text{ m grid_spacing}^{-1}$ (vertical line in b) and $\text{CSI} = 30 \text{ min}$. The GRAD vs. $\ln(\text{CHL})$ contour (b) was conditioned on $\text{SST} = 13.2^\circ\text{C}$ (vertical line in a) and $\text{CSI} = 30 \text{ min}$. The areas without contours indicate no data. The y-axis labels are linear.

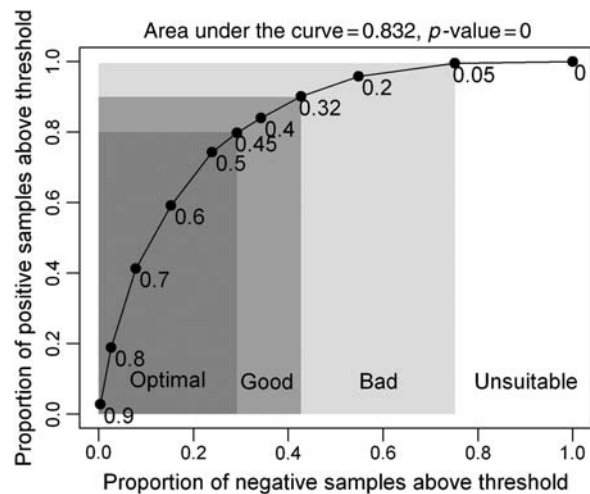


Figure 5. Receiver operating characteristic (ROC) curve for the best model. The proportion of false positives (negative samples classified as positive) are indexed to the x-axis and the proportion of true positives (positive samples classified as positive) to the y-axis for sequential levels of the fitted probability (the values indicated along the ROC curve). The four habitat classes are: optimal, highest probabilities, including 80% of positive samples; good, including 10% of the total positive samples not included in the optimal habitat; bad, including 9% of the total positive samples not included in the optimal and good habitat; and unsuitable, including <1% of the total positive samples not included in the other classes.

sardine habitat in this area. Before this, the few sardine caught in the area were considerably smaller and were presumably non-migrating. Moreover, the biomass density of sardine increases from the initial detection and peaks within 2–4 weeks. A notable exception to this general trend was in spring 2005, when there were large catches of adult sardine in the absence of suitable habitat. However, those fish were considerably smaller than the large migrating sardine and were likely from the very strong year class of 2003, recruited and detected in the area in 2004 as age-0 fish (not shown).

Discussion

During spring, sardine spawn in waters with particular non-linear combinations of SST, $\ln(\text{CHL})$, and GRAD [Equation (1)]. Nevertheless, the values of SST and CHL associated with sardine presence in this study are similar to those described previously using both smaller (Checkley *et al.*, 2000; Reiss *et al.*, 2008) and more comprehensive (Weber and McClatchie, 2010) sets of data collected from ship- and satellite-based sensors. For example, concentrations of sardine eggs have been observed previously associated with SST values of 12–16°C (Checkley *et al.*, 2000; Lynn, 2003; Jacobson *et al.*, 2005; Reiss *et al.*, 2008). Off Oregon, where sardine spawn later in the year, most of the spawning biomass appears to be found where SST is between 14 and 16°C (Emmett *et al.*, 2005). The model presented here, based on sardine egg presence over a 12-year period, demonstrates that spawning sardine are found in water where the SST lies between 11 and 16°C, but mainly between 13 and 14°C.

Sardine are opportunistic feeders, utilizing phytoplankton, microzooplankton, and macrozooplankton (Emmett *et al.*, 2005), with zooplankton comprising most of their carbon intake

(Espinoza *et al.*, 2009). Within the temperature range 12–14.5°C (cf. Figure 3), the model suggests that sardine most often reside in water with CHL values between 0.3 and 1 mg m⁻³, similar to the results of Reiss *et al.* (2008). The model also indicates that sardine avoid areas with high CHL and low SST, indicative of freshly upwelled waters, but can occupy areas with moderate CHL up to >10 mg m⁻³ provided SST is >13°C. The medium-to-low CHL and mild temperature ranges encountered for Pacific sardine eggs may be indicative of curl-driven upwelling areas, with large zooplankton biomasses exerting predation pressure on the phytoplankton community (Ryckaczewski and Checkley, 2008). Such conditions result in smaller phyto- and zooplankton than coastal upwelling regions. Also, because of the lower turbulence in relation to the coastal upwelling areas, curl-driven upwelling regions favour the feeding of sardine larvae (Bakun and Parrish, 1982), and probably adults (Ryckaczewski and Checkley, 2008), promoting the retention of eggs and larvae. Therefore, the apparent preference for these mid-range SST values, which are well within the physiological limits for sardine (Martínez-Porchas *et al.*, 2009), could be related not only to the potentially poor development of sardine larvae at temperatures <13°C (Lasker, 1964), but also to there being inadequate food for adults in freshly upwelled, cooler, coastal waters (Ryckaczewski and Checkley, 2008) and warmer, oligotrophic, oceanic waters (Checkley *et al.*, 2000).

A statistically significant relationship between sardine presence and GRAD is described here and provides a remotely sensed variable that improves the predictions made within the space created by SST and CHL, and possibly constrains the offshore extent of sardine habitat. The habitat model suggests an increase in the probability of a positive sardine encounter when GRAD values are high, indicative of high geostrophic flow, as expected in mesoscale oceanic and coastal eddies (Mantyla *et al.*, 2008). These structures could play an important role in the recruitment of pelagic larvae to adult populations (Loggerwell and Smith, 2001).

The model fit and predictions were not influenced by the size of the SSB. This is because, for the biomass levels considered here, 0.5–1.3 million tonnes (Hill *et al.*, 2009), the area of distribution is largely independent of the biomass (cf. Figure 6 in Barange *et al.*, 2009). Consequently, habitat is the main driver of the large-scale spatial distribution of sardine, and the model predicts the distribution of spawning sardine in the sampling region accurately. Furthermore, the model predicts the presence of spring spawning habitat to the south in coastal waters off Baja California (Figure 7). Baumgartner *et al.* (2008) sampled spawning sardine in that area which were believed to belong to the northern stock.

It was further assumed that the habitat model predicts the potential habitat for non-spawning adult sardine, a hypothesis largely supported by comparing the model predictions with independent information from the fishery in the northeast Pacific and trawl surveys in proximity to the Columbia River mouth, where large catches during summer and the absence of sardine during the rest of the year were accurately represented by the monthly habitat predictions. The analysis of monthly predictions show that sardine habitat recedes south in the winter, supporting historical observations that, in general, sardine migrating from the south in spring do not overwinter in the northeast Pacific (Ware, 1999; Schweigert and McFarlane, 2001). However, smaller sardine including age-0 fish were sampled occasionally, most likely as a consequence of local recruitment events (Emmett *et al.*, 2005). It is unlikely that these smaller fish migrated

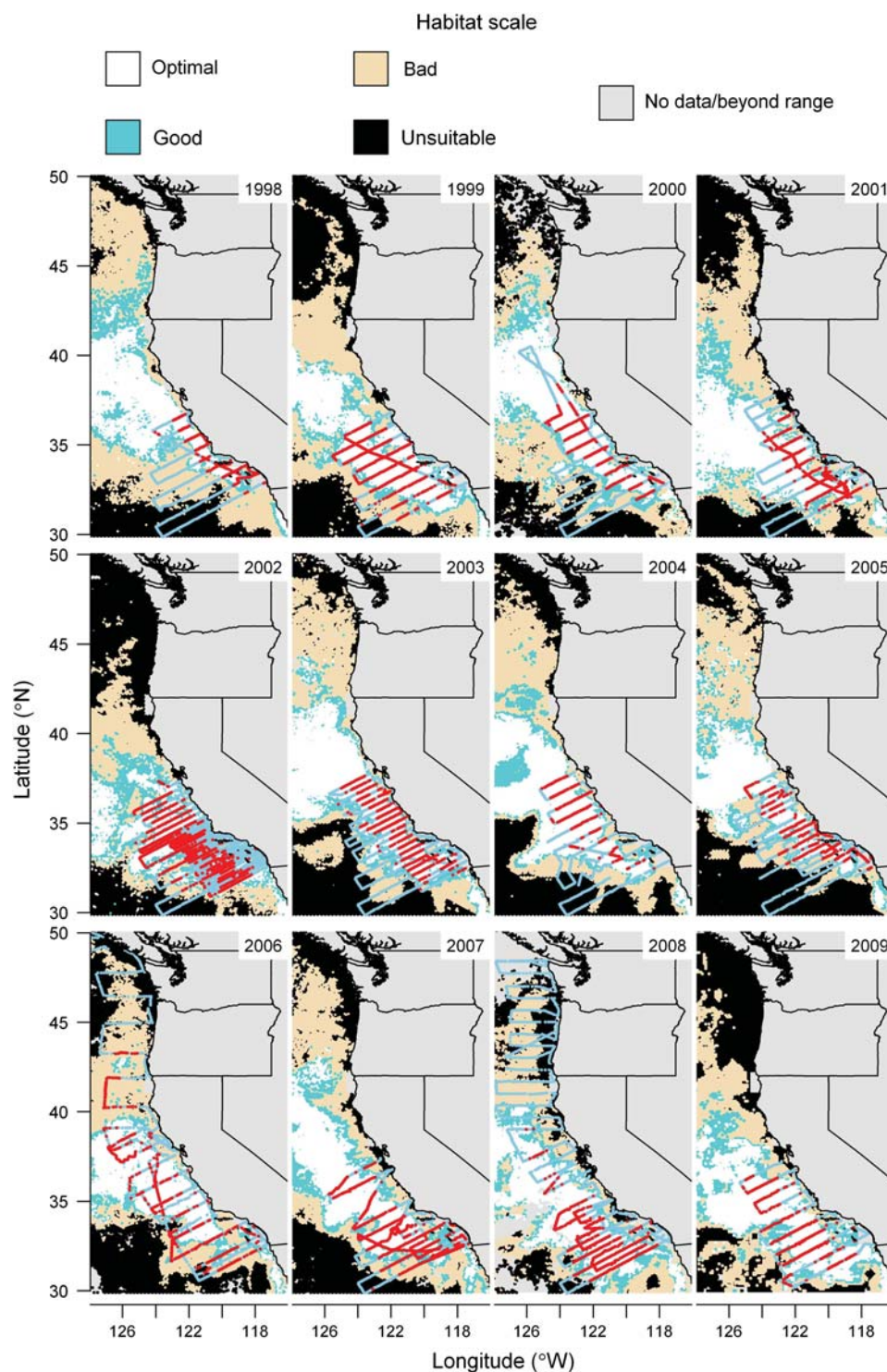


Figure 6. Potential sardine habitat overlaid with the presence (red) and absence (blue) of egg data sampled with a CUFES during spring surveys. Each habitat prediction map was constructed without including in the fitted model data from the respective year.

from the southern spawning grounds (Clark and Janssen, 1945), but instead were likely spawned and remained locally throughout the year. Similar observations have been made in Canadian waters (McFarlane and Beamish, 2001).

The model provides an accurate prediction of sardine habitat and the species' seasonal dynamics. The significant statistical

relationships encountered using a small set of environmental variables summarize ecological mechanisms without requiring their explicit definition (Whittaker *et al.*, 1973). Also, the model shows that a non-linear multivariate combination of environment variables defines species habitat better than a range of environmental variables analysed in isolation or in linear combination.

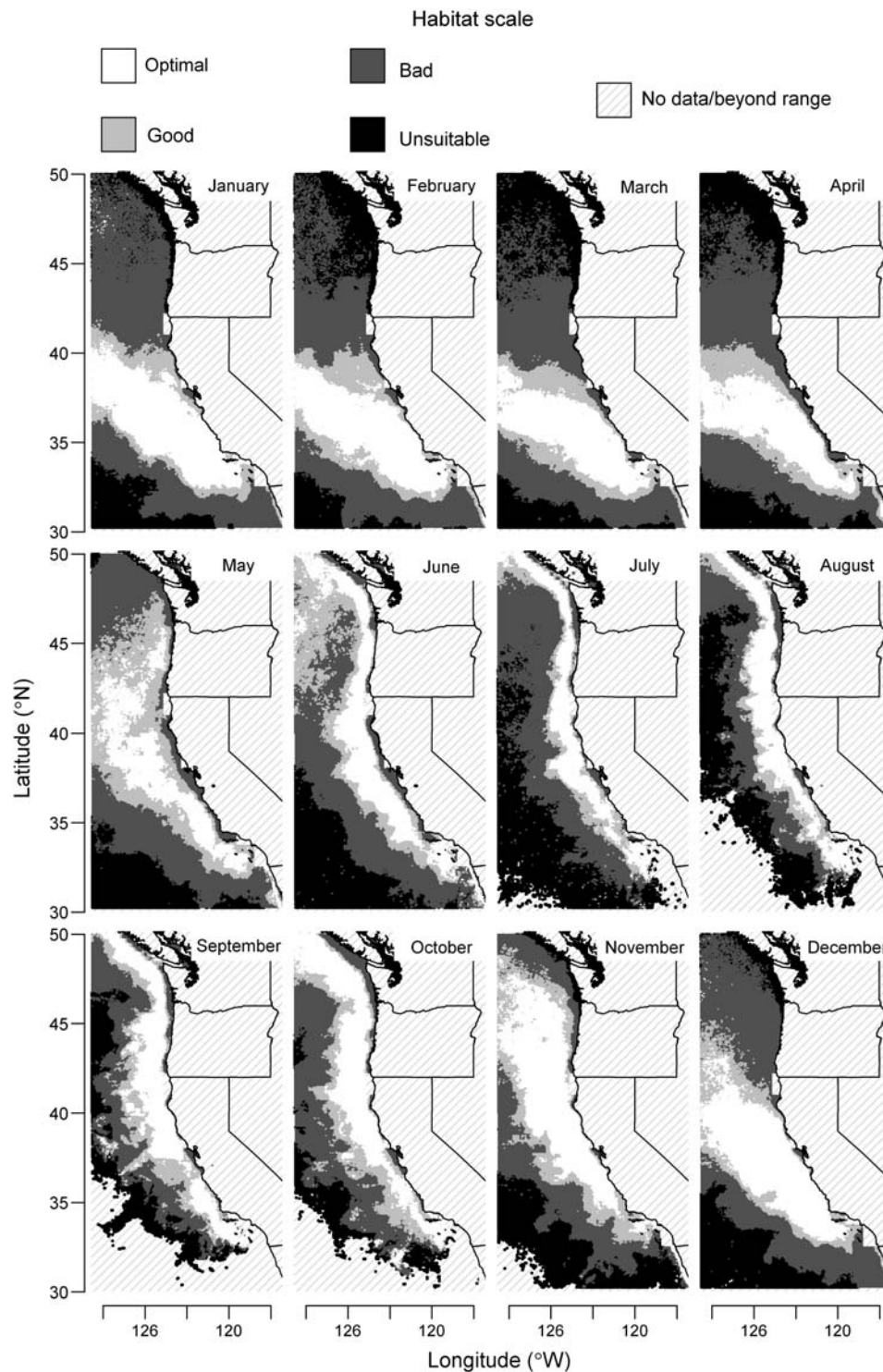


Figure 7. Monthly average sardine habitat, 1998–2009, predicted from the selected GAM [Equation (1)].

This fact reinforces the idea that ecological mechanisms are intrinsically non-linear (Hsieh *et al.*, 2005), a characteristic that applies to both various environment-to-species relationships and their combinations.

The procedure applied here has potential for application to many other migratory epipelagic species in the CC and other marine ecosystems. As demonstrated, data from a variety of large-scale surveys (e.g. Barange *et al.*, 2005; this paper) or

fisheries landings (e.g. Solanki *et al.*, 2010), or both, can be combined with remotely sensed environmental information to construct potential habitat models with temporal and spatial resolutions of approximately days and tens of kilometres. The benefits for such predictive models are many, including optimizing survey-sampling effort, interpreting survey results in the context of potential habitat, predicting or explaining fishing effort, understanding seasonal migrations and trophic-level

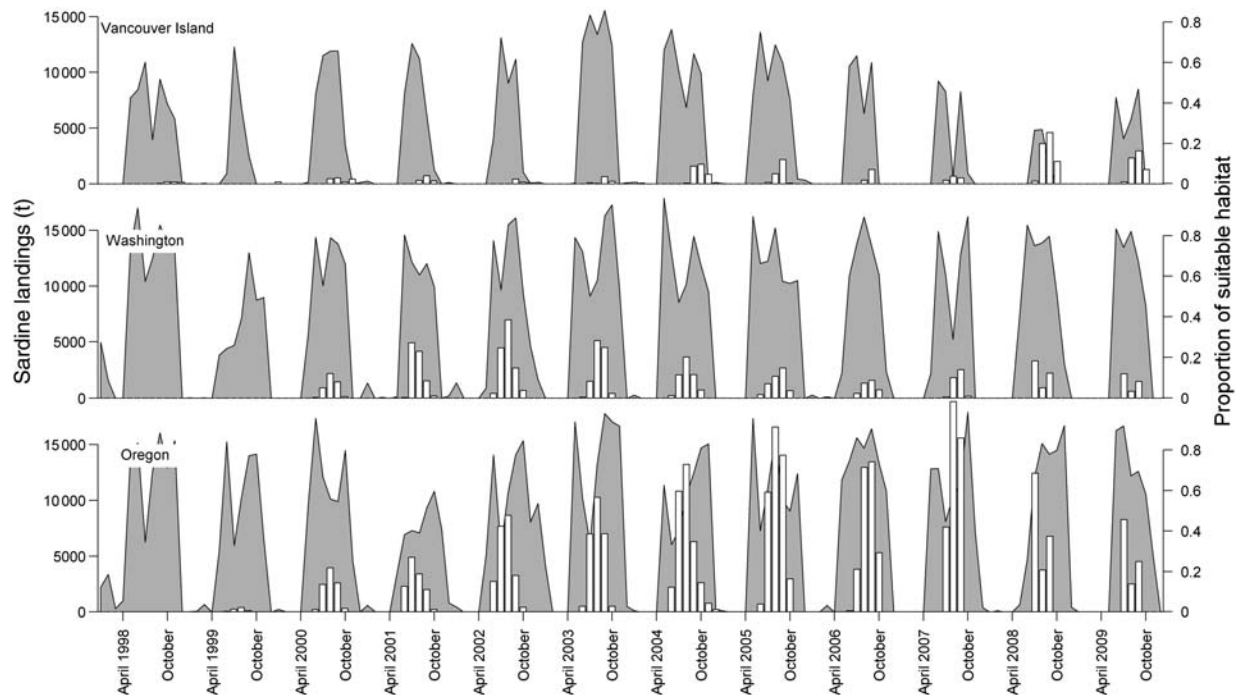


Figure 8. Time-series of potential sardine habitat (grey) and commercial sardine landings (bars) in Oregon, Washington, and south of Vancouver Island. The sardine habitat was evaluated as the proportion of the respective fishing area (polygons centred at $125^{\circ}08.1'W$ $45^{\circ}42.12'N$, $124^{\circ}59.69'W$ $47^{\circ}19.77'N$, and $126^{\circ}49.91'W$ $49^{\circ}23.55'N$, respectively) containing good and optimal habitat.

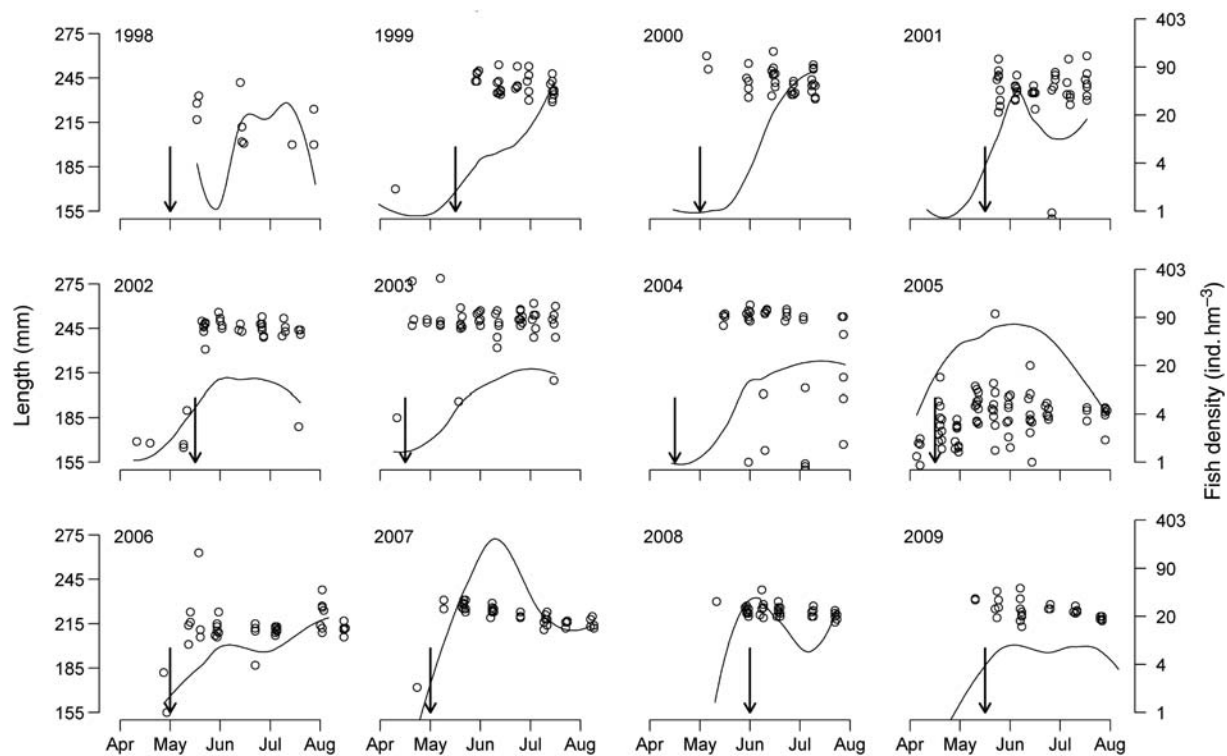


Figure 9. Evolution of adult sardine lengths (circles) and mean biomass densities (local smoother line) inferred from surface trawl catches off the mouth of the Columbia River ($124^{\circ}19.14'W$ $46^{\circ}11.43'N$). The arrows indicate the estimated time of arrival of potential sardine habitat.

interactions, and predicting changes in habitat location and habitat use under scenarios of climate change. If the model fails to predict distributions of sardine, perturbations of the current oceanographic and biological processes summarized by the non-linear combination of SST, $\ln(\text{CHL})$, and GRAD may be indicated. Therefore, the accuracy of model predictions may also serve as an indicator of changes in the state of ecosystem, or so-called regime shifts.

This analysis has provided, for the first time, the environmental basis behind the migration cycle of the northern stock of Pacific sardine reported by Clark and Janssen (1945). The routes taken during the northward and southward migrations appear to differ in their proximity to the coast. During late autumn, the retreat of sardine habitat and the lack of significant landings off Oregon and Washington suggest southward migration mainly offshore, even to the west of 130°W. In contrast, during northward migration, compression of northern cooler waters into the coast gives sardine little option other than to migrate within a coastal filament of habitat. These findings unify an assortment of spatially and temporally scattered observations of sardine, such as spring spawning in the southern California Bight and offshore, the summer fishery of large sardine in the northeast Pacific, the appearance of larger, fatter sardine in central (Clark, 1936) and southern California (Higgins, 1926) a few months after their disappearance from northern waters, and finally their re-aggregation off southern California in spring. This evolution appears to be related to that of the Transition Zone Chlorophyll Front (TZCF; Polovina *et al.*, 2001), perhaps resulting from the same mesoscale processes. The TZCF spans the entire northern Pacific Ocean (Bograd *et al.*, 2004), but runs almost parallel to the coast along the eastern Pacific margin (Polovina *et al.*, 2001). The seasonality of the TZCF appears to regulate the northward and offshore limits of sardine habitat (cf. Figure 2 in Polovina *et al.*, 2001).

The habitat model, including time-invariant relationships and excluding temporal and spatial explanatory variables, consistently explained a large proportion of the variability in egg data and predicted the spatio-temporal evolution of the sardine stock for more than a decade. The model, which is based solely on remotely sensed information, can be used to map, immediately before a survey, the potential habitat of spawning and non-spawning adult sardine from the northern stock. The predictions of potential sardine habitat reflect the most likely area of sardine presence and, although the proportion of the habitat utilized by sardine varies from survey to survey as a consequence of their aggregating behaviour, substantial gains in survey efficiencies were shown. Oceanographic changes within the environmental window (7–14 d) of each composite satellite image, especially offshore, did not show significant discrepancies with *in situ* measurements, suggesting that small spatial- and temporal-scale variability has a negligible effect on short-term forecasting. Therefore, potential sardine habitat can be predicted with the latest available satellite data for planning CUFES, acoustic, and trawl surveys with gains as high as half the duration of past surveys. These saved resources can be reallocated to sample areas of potential sardine habitat better, perhaps reducing sampling bias attributable to incomplete sampling of the entire population. Moreover, the sampling variance (Aglen, 1983), which is often the greatest source of random error in acoustic surveys (Demer, 2004), would be reduced. The model predictions suggest that acoustic-trawl surveys conducted during summer may be more efficient than

similar surveys conducted in spring, because the offshore limits of potential sardine habitat are well marked and close to shore. Also, the nearshore operations of the fishing fleet might create synergies for the collection of biological samples. Moreover, the longer periods of daylight would provide more acoustic sampling, which is ineffective when sardine migrate to the surface at night, and hence a reduction in sampling variance.

Supplementary data

Supplementary data are available at the ICES/JMS online version of the manuscript.

Acknowledgements

We thank the scientists and technicians from NOAA and SIO that collected and archived the data from the DEPM and CalCOFI surveys, in particular Dave Griffith, Amy Hays and Susie Jacobson, and Kevin Hill for providing sardine landing data. We are also indebted to the various institutions responsible for collecting and disseminating the remotely sensed information from AVISO, Pathfinder, MODIS, and SeaWiFS, and to the staff of CoastWatch West Coast Regional Node, particularly Dave Foley, for providing the data products and related support. We also thank Vardis Tsontos for his guidance through the many portals of remotely sensed information, and to our colleagues at the Southwest Fisheries Science Center (SWSFC) for constructive input. We thank the R Development Core Team and all the R contributors for maintaining the software that facilitated the GAMs and remaining analysis, and Russ Vetter, Director of the Fisheries Resources Division at SWFSC, and the Portuguese Foundation for Science and Technology (FCT-MCES) for funding JPZ's post-doctoral internship (SFRH/BPD/44834/2008) with DAD at SWFSC. Finally, we thank Arnaud Bertrand and anonymous reviewers for the positive input they made that improved the manuscript substantially.

References

- Aglen, A. 1983. Random errors of acoustic fish abundance estimates in relation to the survey grid density applied. *FAO Fisheries Report*, 300: 293–298.
- Bakun, A., and Parrish, R. H. 1982. Turbulence, transport, and pelagic fish in the California and Peru Current systems. *CalCOFI Reports*, 23: 99–112.
- Barange, M., Coetzee, J., Takasuka, A., Hill, K., Gutierrez, M., Oozeki, Y., van der Linden, C., *et al.* 2009. Habitat expansion and contraction in anchovy and sardine populations. *Progress in Oceanography*, 83: 251–260.
- Barange, M., Coetzee, J., and Twatwa, N. 2005. Strategies of space occupation in anchovy and sardine in the southern Benguela: the role of stock size and intra-specific competition. *ICES Journal of Marine Science*, 62: 645–654.
- Barry, S., and Elith, J. 2006. Error and uncertainty in habitat models. *Journal of Applied Ecology*, 43: 413–423.
- Baumgartner, T., Durazo, R., Lavaniegos, B., Gaxiola, G., Gómez, J., and García, J. 2008. Ten years of change from IMECOCAL observations in the southern region of the California Current ecosystem. *GLOBEC International Newsletter*, 14: 43–54.
- Bograd, S. J., Foley, D. G., Schwing, F. B., Wilson, C., Laurs, R. M., Polovina, J. J., Howell, E. A., *et al.* 2004. On the seasonal and inter-annual migrations of the transition zone chlorophyll front. *Geophysical Research Letters*, 31: L17204.
- Checkley, D. M., Dotson, R. C., and Griffith, D. A. 2000. Continuous, underway sampling of eggs of Pacific sardine (*Sardinops sagax*) and

- northern anchovy (*Engraulis mordax*) in spring 1996 and 1997 off southern and central California. Deep Sea Research II, 47: 1139–1155.
- Chilès, J.-P., and Delfiner, P. 1999. Geostatistics—Modelling Spatial Uncertainty. Wiley Series in Probability and Statistics. John Wiley, New York. 695 pp.
- Clark, F. N. 1936. Interseasonal and intraseasonal changes in size of the California sardine (*Sardinops caerulea*). California Department of Fish and Game Fish Bulletin, 47: 31 pp.
- Clark, F. N., and Janssen, J. F. 1945. Movements and abundance of the sardine as measured by tag returns. California Department of Fish and Game Fish Bulletin, 61: 7–42.
- Curtis, K. A. 2004. Fine scale spatial pattern of Pacific sardine (*Sardinops sagax*) and northern anchovy *Engraulis mordax* eggs. Fisheries Oceanography, 13: 239–254.
- Demer, D. A. 2004. An estimate of error for the CCAMLR 2000 survey estimate of krill biomass. Deep Sea Research II, 51: 1237–1251.
- Diniz-Filho, J. A. F., Bini, M. L., and Hawkins, B. A. 2003. Spatial autocorrelation and red herrings in geographical ecology. Global Ecology and Biogeography, 12: 53–64.
- Emmett, R. L., Brodeur, R. D., Miller, T. W., Pool, S. S., Bentley, P. J., Krutzikowsky, G. K., and McCrae, J. 2005. Pacific sardine (*Sardinops sagax*) abundance, distribution and ecological relationships in the Pacific Northwest. CalCOFI Reports, 46: 122–143.
- Espinoza, P., Bertrand, A., van der Lingen, C. D., Garrido, S., and de Mendiola, B. R. 2009. Diet of sardine (*Sardinops sagax*) in the northern Humboldt Current system and comparison with the diets of clupeoids in this and other eastern boundary upwelling systems. Progress in Oceanography, 83: 242–250.
- Fawcett, T. 2006. An introduction to ROC analysis. Pattern Recognition Letters, 27: 861–874.
- Guisan, A., and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. Ecological Modelling, 135: 147–186.
- Hanley, J. A., and McNeil, B. J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143: 29–36.
- Higgins, E. 1926. A study of fluctuations in the sardine fishery at San Pedro. California Department of Fish and Game Fish Bulletin, 11: 125–158.
- Hill, K. T., Lo, N. C. H., Macewicz, B. J., Crone, P. R., and Félix-Uraga, R. 2009. Assessment of the Pacific sardine resource in 2009 for US management in 2010. NOAA Technical Memorandum, NOAA-TM-NMFS-SWFSC-452. 156 pp.
- Hill, K. T., Lo, N. C. H., Macewicz, B. J., and Félix-Uraga, R. 2006. Assessment of the Pacific sardine (*Sardinops sagax caerulea*) population for US management in 2007. NOAA Technical Memorandum, NOAA-TM-NMFS-SWFSC-396. 79 pp.
- Hsieh, C., Glaser, S., Lucas, A. J., and Sugihara, G. 2005. Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. Nature, 435: 336–340.
- Jacobson, L. D., Bograd, S. J., Parrish, R. H., Mendelssohn, R., and Schwing, F. B. 2005. An ecosystem-based hypothesis for climatic effects on surplus production in California sardine (*Sardinops sagax*) and environmentally dependent surplus production models. Canadian Journal of Fisheries and Aquatic Sciences, 62: 1782–1796.
- Lasker, R. 1964. An experimental study of the effect of temperature on incubation time, development and growth of Pacific sardine embryos and larvae. Copeia, 2: 399–406.
- Lluch-Belda, D., Lluch-Cota, D. B., Hernández-Vázquez, S., Salinas-Zavala, C., and Schwartzlose, R. A. 1991. Sardine and anchovy spawning as related to temperature and upwelling in the California Current system. CalCOFI Reports, 32: 105–111.
- Lo, N. C. H., Macewicz, B. J., Griffith, D. A., and Charter, R. L. 2007. Spawning biomass of Pacific sardine (*Sardinops sagax*) off US and Canada in 2006. NOAA Technical Memorandum, NOAA-TM-NMFS-SWFSC-401. 32 pp.
- Loggerwell, E. A., and Smith, P. E. 2001. Mesoscale eddies and survival of late stage Pacific sardine (*Sardinops sagax*) larvae. Fisheries Oceanography, 10: 13–25.
- Lynn, R. J. 2003. Variability in the spawning habitat of Pacific sardine (*Sardinops sagax*) off southern and central California. Fisheries Oceanography, 12: 541–553.
- MacCall, A. 2009. Mechanisms of low-frequency fluctuation in sardine and anchovy populations. In Climate Change and Small Pelagic Fish, pp. 285–299. Ed. by D. Checkley, J. Alheit, Y. Oozeki, and C. Roy. Cambridge University Press, Cambridge, UK. 372 pp.
- Macewicz, B. J., and Abramenkoff, D. N. 1993. Collection of jack mackerel, *Trachurus symmetricus*, off southern California during 1991 cooperative US–USSR cruise. SWFSC–NMFS Administrative Report, LJ-93-07. 13 pp.
- Mantyla, A. W., Bograd, S. J., and Venrick, E. L. 2008. Patterns of chlorophyll *a* and primary productivity cycles in the southern California Bight. Journal of Marine Systems, 73: 48–60.
- Martínez-Porchas, M., Hernández-Rodríguez, M., and Buckle-Ramírez, L. F. 2009. Thermal behavior of the Pacific sardine (*Sardinops sagax*) acclimated to different thermal cycles. Journal of Thermal Biology, 34: 372–376.
- McClatchie, S. 2009. Report on the NMFS California Current Ecosystem Survey (CCES) (April and July–August 2008). NOAA Technical Memorandum, NOAA-TM-NMFS-SWFSC-438. 98 pp.
- McCullagh, P., and Nelder, J. A. 1997. Generalized Linear Models, 2nd edn. Chapman and Hall, London. 511 pp.
- McFarlane, G. A., and Beamish, R. J. 2001. The re-occurrence of sardines off British Columbia characterizes the dynamic nature of regimes. Progress in Oceanography, 49: 151–165.
- McFarlane, G. A., Schweigert, J., MacDougall, L., and Hrabok, C. 2005. Distribution and biology of Pacific sardine (*Sardinops sagax*) off British Columbia, Canada. CalCOFI Reports, 46: 144–160.
- Polovina, J. J., Howell, E., Kobayashi, D. R., and Seki, M. P. 2001. The Transition Chlorophyll Front, a dynamic global feature defining migration and forage habitat for marine resources. Progress in Oceanography, 49: 469–483.
- R Development Core Team. 2009. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>.
- Radovich, J. 1982. The collapse of the California sardine fishery: what have we learned. CalCOFI Reports, 23: 56–78.
- Reiss, C. S., Checkley, D. M., and Bograd, S. J. 2008. Remotely sensed spawning habitat of Pacific sardine (*Sardinops sagax*) and northern anchovy (*Engraulis mordax*) within the California Current. Fisheries Oceanography, 17: 126–136.
- Ryckaczewski, R. R., and Checkley, D. M. 2008. Influence of ocean winds on the pelagic ecosystem in upwelling regions. Proceedings of the National Academy of Sciences of the USA, 105: 1965–1970.
- Schweigert, J., and McFarlane, G. A. 2001. Stock assessment and recommended harvest for Pacific sardine in 2002. Canadian Science Advisory Secretariat Research Document, 2001/126.
- Smith, P. E. 1978. Precision of sonar mapping for pelagic fish assessment in the California Current. Journal du Conseil International pour l'Exploration de la Mer, 38: 33–40.
- Smith, P. E. 2005. A history of proposals for subpopulation structure in the Pacific sardine (*Sardinops sagax*) population off western North America. CalCOFI Report, 46: 75–82.
- Smith, P. E., and Moser, H. G. 2003. Long-term trends and variability in the larvae of Pacific sardine and associated fish species of the California Current region. Deep Sea Research II, 50: 2519–2536.
- Solanki, H. U., Prakash, P., Dwivedi, R. M., Nayak, S., Kulkarni, A., and Somvamshi, V. S. 2010. Synergistic application of oceanographic variables from multi-satellite sensors for forecasting potential fishing zones: methodology and validation results. Remote Sensing, 31: 775–789.

- Ware, D. M. 1999. Life history of Pacific sardine and a suggested framework for determining a BC catch quota. Canadian Stock Assessment Secretariat Research Document, 99/204. 19 pp.
- Weber, E. D., and McClatchie, S. 2010. Predictive models of northern anchovy *Engraulis mordax* and Pacific sardine *Sardinops sagax* spawning habitat in the California Current. Marine Ecology Progress Series, 406: 251–263.
- Whittaker, R. H., Levin, S. A., and Root, R. B. 1973. Niche, habitat and ecotope. The American Naturalist, 107: 321–338.
- Wood, S. 2006. Generalized Additive Models: an Introduction with R. Texts in Statistical Sciences. Chapman and Hall/CRC, Boca Raton, FL. 391 pp.
- Wood, S. N., and Augustin, N. H. 2002. GAMS with integrated model selection using penalized regression splines and applications to environmental modeling. Ecological Modeling, 157: 157–177.
- Zwolinski, J., Mason, E., Oliveira, P. B., and Stratoudakis, Y. 2006. Fine-scale distribution of sardine (*Sardina pilchardus*) eggs and adults during a spawning event. Journal of Sea Research, 56: 294–304.